

**Title:** Ancient genomes revisit the ancestry of domestic and Przewalski's horses

**Authors:** Charleen Gaultz<sup>1,\*</sup>, Antoine Fages<sup>1,2,\*</sup>, Kristian Hanghøj<sup>1,2</sup>, Anders Albrechtsen<sup>3</sup>, Naveed Khan<sup>1,4</sup>, Mikkel Schubert<sup>1</sup>, Andaine Seguin-Orlando<sup>1,5</sup>, Ivy J. Owens<sup>6,7</sup>, Sabine Felkel<sup>8</sup>, Olivier Bignon-Lau<sup>9</sup>, Peter de Barros Damgaard<sup>1</sup>, Alissa Mittnik<sup>10</sup>, Azadeh Mohaseb<sup>11</sup>, Hossein Davoudi<sup>12</sup>, Saleh Alquraishi<sup>13</sup>, Ahmed H. Alfarhan<sup>13</sup>, Khaled A. S. Al-Rasheid<sup>13</sup>, Eric Crubézy<sup>2</sup>, Norbert Benecke<sup>14</sup>, Sandra Olsen<sup>15</sup>, Dorcas Brown<sup>16</sup>, David Anthony<sup>16</sup>, Ken Massy<sup>17</sup>, Vladimir Pitulko<sup>18</sup>, Aleksei Kasparov<sup>18</sup>, Gottfried Brem<sup>8</sup>, Michael Hofreiter<sup>19</sup>, Gulmira Mukhtarova<sup>20</sup>, Nurbol Baimukhanov<sup>21</sup>, Lembi Lõugas<sup>22</sup>, Vedat Onar<sup>23</sup>, Philipp W. Stockhammer<sup>10,17</sup>, Johannes Krause<sup>10</sup>, Bazartseren Boldgiv<sup>24</sup>, Sainbileg Undrakhbold<sup>24</sup>, Diimaajav Erdenebaatar<sup>25</sup>, Sébastien Lepetz<sup>11</sup>, Marjan Mashkour<sup>11</sup>, Arne Ludwig<sup>26</sup>, Barbara Wallner<sup>8</sup>, Victor Merz<sup>27</sup>, Ilja Merz<sup>27</sup>, Viktor Zaibert<sup>28</sup>, Eske Willerslev<sup>1</sup>, Pablo Librado<sup>1</sup>, Alan K. Outram<sup>6,+</sup>, Ludovic Orlando<sup>1,2,+</sup>.

**Affiliations:**

<sup>1</sup>Centre for GeoGenetics, Natural History Museum of Denmark, Øster Voldgade 5-7, 1350K Copenhagen, Denmark

<sup>2</sup>Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse UMR 5288, Université de Toulouse, CNRS, Université Paul Sabatier, France

<sup>3</sup>Bioinformatics Center, Department of Biology, University of Copenhagen, 2200N Copenhagen, Denmark

<sup>4</sup>Department of Biotechnology, Abdul Wali Khan University, Mardan, Pakistan

<sup>5</sup>National High-Throughput DNA Sequencing Center, Copenhagen, Denmark

<sup>6</sup>Department of Archaeology, University of Exeter, Exeter EX4 4QE, UK

<sup>7</sup>The Charles McBurney Laboratory for Geoarchaeology, Division of Archaeology, University of Cambridge, Cambridge CB2 3DZ, UK

<sup>8</sup>Institute of Animal Breeding and Genetics, University of Veterinary Medicine Vienna, 1210 Vienna, Austria

<sup>9</sup>Équipe Ethnologie préhistorique, ArScAn, CNRS UMR 7041, MAE René-Ginouvès, 92023 Nanterre Cedex, France

<sup>10</sup>Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Kahlaische Str. 10, 07745 Jena, Germany

<sup>11</sup>Centre National de la Recherche Scientifique, Muséum National d'Histoire Naturelle, Archéozoologie, Archéobotanique, Sociétés, Pratiques et Environnements (UMR 7209), 55 rue Buffon, 75005 Paris, France

<sup>12</sup>Department of Archaeology, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran

<sup>13</sup>Zoology Department, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

<sup>14</sup>German Archaeological Institute, Department of Natural Sciences, Berlin, 14195 Berlin, Germany

<sup>15</sup>Biodiversity Institute/Museum of Natural History, University of Kansas, Spooner Hall, 1340 Jayhawk Blvd., Lawrence, KS, 66045 USA

<sup>16</sup>Anthropology Department, Hartwick College, Oneonta, NY 13820

<sup>17</sup>Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, LudwigMaximilians-University Munich, Schellingstraße 12, 80799 München, Germany

<sup>18</sup>Institute for the History of Material Culture, Russian Academy of Sciences, 18 Dvortsovaya nab., St. Petersburg, 191186, Russia

<sup>19</sup>University of Potsdam, Faculty of Mathematics and Natural Sciences, Institute for Biochemistry and Biology, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany

<sup>20</sup>State historical-cultural reserve-museum «Issyk», 040400 Almaty region, Enbekshikazak distric, Kazakhstan

<sup>21</sup>Shejire DNA project, Abai ave. 150/230, 050046 Almaty, Kazakhstan

<sup>22</sup>Archaeological Research Collection, Tallinn University, Rütli 10, 10130 Tallinn, Estonia

<sup>23</sup>Osteoarchaeology Practice and Research Center & Department of Anatomy, Faculty of Veterinary Medicine, Istanbul University, 34320, Avcılar, Istanbul, Turkey

<sup>24</sup>Ecology Group, Department of Biology, School of Arts and Sciences, National University of Mongolia, Ulaanbaatar 14201, Mongolia

<sup>25</sup>Department of Archaeology, Ulaanbaatar State University, Ulaanbaatar 51, Mongolia

<sup>26</sup>Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Berlin 10315, Germany

<sup>27</sup>S.Toraighyrov Pavlodar State University, Joint Research Center for Archeological Studies named after A.Kh. Margulan, Lomov St. 64, cab. 102, Pavlodar, Kazakhstan

<sup>28</sup>Scientific Centre of Archaeology and Ethnology, Kostanay State Pedagogical Institute, 110000, Kostanay, Tauelsizdik street, 118, Kazakhstan

\*These authors contributed equally to this work. +Corresponding authors: Alan Outram, [A.K.Outram@exeter.ac.uk](mailto:A.K.Outram@exeter.ac.uk); Ludovic Orlando, [ludovic.orlando@univ-tlse3.fr](mailto:ludovic.orlando@univ-tlse3.fr)

**Abstract** (116 words) The Eneolithic Botai culture of the Central Asian steppes provides the earliest archaeological evidence for horse husbandry, ~ 5,500 ya, but the exact nature of early horse domestication remains controversial. We generated 42 ancient horse genomes, including 20 from Botai. Compared to 46 published ancient and modern horse genomes, our data indicate that Przewalski's horses are the feral descendants of horses herded at Botai and not truly wild horses. All domestic horses dated from ~4,000 ya to present only show ~2.7% of Botai-related ancestry. This indicates that a massive genomic turnover underpins the expansion of the horse stock that gave rise to modern domesticates, which coincides with large-scale human population expansions during the Early Bronze Age.

**One Sentence Summary** (103 characters) Earliest herded horses were the ancestors of feral Przewalski's horses but not of modern domesticates.

**Main text:** Horses revolutionized human mobility, economy, and warfare (1). They are also associated with the spread of Indo-European languages (2), new forms of metallurgy (3) and provided the fastest land transport until modern times. Together with the lack of diachronic changes in horse morphology (4) and herd structure (5, 6), the scarce archaeological record hampered the study of early domestication. With their preponderance of horse remains, Eneolithic sites (5<sup>th</sup> and 4<sup>th</sup> Mill BCE) of the Pontic-Caspian steppe (2, 7) and the northern steppe of Kazakhstan (6, 8) have attracted the most attention.

We reconstructed the phylogenetic origins of the Eneolithic horses associated with the Botai culture of northern Kazakhstan, representing the earliest domestic horses (6, 8). This culture was characterized by a sudden shift from mixed hunting/gathering to an extreme focus on horses, and larger, more sedentary settlements (5). Horse dung on site (6), as well as evidence for poleaxing and against selective body part transportation, suggest controlled slaughter at settlements rather than hunting (9). Tools associated with leather thong production, bit-related dental pathologies (7,10) and equine milk fats within ceramics support pastoral husbandry, involving milking and harnessing (8).



Geological surveys at the Botai culture site of Krasnyi Yar, Kazakhstan described a polygonal enclosure of ~20 x 15 meters with elevated phosphorus and sodium levels (6), likely corresponding to a horse corral. We revealed a similar enclosure at the eponymous Botai site, ~100 kilometers west of Krasnyi Yar (Fig. 1A), showing close-set post molds, merging to form a palisade trench, and a line of smaller parallel postholes inside (Fig. 1B). Radiocarbon dates on horse bones from these postholes are consistent with the Botai culture (11). The presence of enclosures at Krasnyi Yar and Botai adds on the evidence supporting horse husbandry.

We sequenced the genomes of 20 horses from Botai, and 22 from across Eurasia and spanning the last ~5,000 years (Table S1). With the published genomes of 18 ancient and 28 modern horses, this provided a comparative panel of three wild archaic horses (~42,800-5,100 ya), seven Przewalski's horses (PH, six modern and one from the 19<sup>th</sup> century), and 78 domesticates (25 Eneolithic, including five from Borly4, Kazakhstan ~5,000 ya; seven Bronze Age ~4,100-3,000 ya; 18 Iron Age ~2,800-2,200 ya; one Parthian and two Roman ~2,000-1,600 ya; three post-Roman ~1,200-100 ya, and 22 modern from 18 breeds).

The 42 ancient genomes, belonging to 31 horse stallions and 11 mares, were sequenced to an average depth-of-coverage of ~1.1-9.3X (median=3.0X). Damage patterns indicative of ancient DNA were recovered (Figs. S8 and S9). Base quality rescaling and termini trimming resulted in average error rates of 0.07%-0.14% per site (Tables S13 and S14).

Principal Component Analysis (PCA) revealed PH and the archaic horses as two independent clusters (Fig. 2A). Within domesticates, all 25 Botai/Borly4 Eneolithic specimens grouped together to the exclusion of all remaining horses.

Phylogenetic reconstruction confirmed that domestic horses do not form a single monophyletic group as expected if descending from Botai (Fig. 2B). Instead, PH form a highly-drifted, monophyletic group, unambiguously nested within Botai/Borly4 horses. All remaining domesticates cluster within a second,

highly-supported monophyletic group (DOM2). Applying TreeMix (12) to the 60 genomes with minimal 3.0X average depth-of-coverage confirmed this tree topology (Fig. S23).

$f_3$ -outgroup and D- statistics (13) support PH as genetically closer to Botai/Borly4 individuals than any DOM2 member (Fig. 2C, Fig. S25 and S26). Finally, ancestry tests (14) confirmed Botai horses as the direct ancestors of Borly4 horses, and the latter as ancestral to the only PH in our dataset pre-dating their massive demographic collapse and introgression of modern domestic genes (15).

$f_3$ -outgroup and D-statistics also revealed that Dunaújváros\_Duk2 (Duk2), the earliest and most basal specimen within DOM2, was divergent to all other DOM2 members. This is not due to sequencing errors since the internal branch splitting from Duk2 and leading to the ancestor of all remaining DOM2 horses is long (Fig. 2B). This suggests instead shared ancestry between Duk2 and a divergent ghost population. We thus excluded Duk2 in admixture graph reconstructions (16) to avoid bias due to contributions from unsampled lineages (Fig. 3).

In the absence of admixture, the best admixture graph matched the trees reconstructed above. We also reconstructed admixture graphs for five additional scenarios with one or two admixture event(s), including between PH and domesticates (15). Bayes Factors best supported a horse domestication history where a first lineage gave rise to Botai/Borly4 and PH horses, while a second lineage founded DOM2 and provided the source of domestic horses during at least the last ~4,000 years, with minimal contribution from the Botai/Borly4 lineage (95% CI=2.0-3.8%).

The limited Botai/Borly4 ancestry amongst DOM2 members concurs with slightly significant negative D-statistics in the form of (((DOM2\_ancient,DOM2\_modern),Botai/Borly4),donkey) for some DOM2 members, spanning a large geographical (Western Europe, Turkey, Iran and Central Asia) and temporal range (from ~3,318 ya to ~1,143 ya; Fig. S28). This suggests sporadic introgression of Botai ancestry into multiple DOM2 herds until the last thousand years. This gene flow was mediated through females since 15 Botai/Borly4 individuals carried mitochondrial haplotypes characteristic of DOM2 matriline (Fig.

S12-S13) but also through males given the persistence of Botai/Borly4-related patriline within DOM2 (Fig. S15-18).

PH are considered as the last true remaining wild horses, that have never been domesticated (15). Our results reveal that they represent instead the feral descendants of horses first herded at Botai. Instead, it appears that their feralization likely involved multiple biological changes.

Metacarpal measurements in 263 ancient and 112 modern horses indicate that PH have become less robust than their Botai/Borly4 ancestors (Fig. 4A). One Botai individual likely showed limited unpigmented areas and leopard spots as it was heterozygous for four mutations at the *TRPM1* locus associated with leopard spotting and carried the ancestral allele at the *PATN1* modifier (17, 18) (Fig. 4B). Individuals homozygous for *TRPM1* mutations are generally almost completely unpigmented and develop congenital stationary night blindness (17). First maintained at Botai by human management, the haplotype associated with leopard spotting was likely selected against and lost once returning wild, leading to the characteristic PH Dun dilution coloration (19). Genomic regions with signatures of positive selection along the phylogenetic branch separating Borly4 and PH showed functional enrichment for genes associated in humans with cardiomyopathies ( $p\text{-values} \leq 0.0496$ ), melanosis and hyperpigmentation (0.0468), and skeletal abnormalities (0.0594) (Table S18), suggesting that at least some of the morpho-anatomical changes associated with feralization were adaptive.

Additionally, significantly negative D-statistics in the form of (((DOM2,PH),archaic),donkey) previously suggested that the extinct, archaic lineage formed by ~5.1-42.7 ky-old horses from Taymyr and Yakutia, contributed to the genetic ancestry of modern domesticates (20, 21). Although we could confirm such D-statistics (Fig. S29), almost all other D-statistics in the form of (((DOM2,Botai/Borly4),archaic),donkey) were not different from zero (Fig. S30). This indicates selection against the archaic ancestry between ~4,977 and ~118 ya (the time interval separating the youngest Borly4 individual and the earliest PH sequenced). Alternatively, the PH lineage admixed with a divergent population of horses, both unrelated

to the archaic lineage and the ghost population that contributed ancestry to Duk2, since D-statistics revealed Duk2 as closer to Borly4 than to PH (Fig. S31).

Lastly, although the genetic load of PH and Botai/Borly4 genomes was equivalent until ~118 ya, it drastically increased in modern animals (Fig. 4C). This accumulation of deleterious variants was thus not associated with PH feralization but with the recent introgression of deleterious variants from modern domesticates and demographic collapse, which hampered purifying selection.

That none of the domesticates sampled in the last ~4,000 years descend from the horses first herded at Botai entails another major implication. It suggests that during the 3<sup>rd</sup> Mill BCE at the latest, another unrelated group of horses became the source of all domestic populations that expanded thereafter. This is compatible with two scenarios. First, Botai-type horses experienced massive introgression capture (22) from a population of wild horses until the Botai ancestry was almost completely replaced. Alternatively, horses were successfully domesticated in a second domestication center and incorporated minute amounts of Botai ancestry during their expansion. We cannot identify the locus of this hypothetical center due to a temporal gap in our dataset throughout the 3<sup>rd</sup> Mill BCE. However, that the DOM2 earliest member was excavated in Hungary adds Eastern Europe to other candidates already suggested, including the Pontic-Caspian steppe (2), Eastern Anatolia (23), Iberia (24), Western Iran and the Levant (25). Notwithstanding the process underlying the genomic turnover observed, the clustering of ~4,023-3,574 year-old specimens from Russia, Romania and Georgia within DOM2 suggests that this clade already expanded throughout the steppes and Europe at the transition between the 3<sup>rd</sup> and 2<sup>nd</sup> Mill BCE, in line with the demographic expansion at ~4,500 ya recovered in mitochondrial Bayesian Skylines (Fig. S14).

This study shows that the horses exploited by the Botai people later became the feral PH. Early domestication most likely followed the ‘prey pathway’ whereby a hunting relationship was intensified until reaching concern for future progeny through husbandry, exploitation of milk and harnessing (7). Other horses, however, were the main source of domestic stock over the last ~4,000 years or more.

Ancient human genomics (26) has revealed considerable human migrations ~5,000 ya involving “Yamnaya” culture pastoralists of the Pontic-Caspian steppe. This expansion might be associated with the genomic turnover identified in horses, especially if Botai horses were best suited to localized pastoral activity than to long distance travel and warfare. Future work must focus on identifying the main source of the domestic horse stock and investigating how the multiple human cultures managed the available genetic variation to forge the many horse types known in history.

## References and notes

1. P. Kelekna, Others, *The horse in human history* (Cambridge University Press New York, 2009).
2. D. W. Anthony, D. Ringe, The Indo-European Homeland from Linguistic and Archaeological Perspectives. *Annu. Rev. Linguist.* **1**, 199–219 (2015).
3. D. W. Anthony, *The Horse, the Wheel, and Language: How Bronze-Age Steppes Shaped the Modern World* (2007).
4. V. Eisenmann, M. Mashkour, in *Les équidés dans le monde méditerranéen antique*, A. Gardeisen, Ed. (Éd. de l'Association pour le développement de l'archéologie en Languedoc-Rousillon, 2005), pp. 41–49.
5. S. L. Olsen, “Early horse domestication on the Eurasian steppe” in *Documenting domestication: New genetic and archaeological paradigms*, M.A.Zeder, D.G.Bradley, E. Emshwiller, B.D. Smith, Eds. (Univ. of California Press, Berkley, 2006), pp. 245–269.
6. S. L. Olsen, Early horse domestication: weighing the evidence. *BAR Int. Series.* **1560**, 81 (2006).
7. D. W. Anthony, D. R. Brown, The Secondary Products Revolution, Horse-Riding, and Mounted Warfare. *J. World. Prehist.* **24**, 131 (2011).

8. A. K. Outram *et al.*, The earliest horse harnessing and milking. *Science*. **323**, 1332–1335 (2009).
9. N. Benecke, A. von den Driesch, “Horse exploitation in the Kazakh steppes during the Eneolithic and Bronze Age” in *Prehistoric steppe adaptation and the horse*, M. Levine, C. Renfrez, K. Boyle, Eds. (McDonald Institute for Archaeological Research, Cambridge, 2003), pp. 69–82.
10. D. Brown, D. Anthony, Bit wear, horseback riding and the Botai site in Kazakhstan. *J. Arc. Sci.* **25**, 331–347 (1998).
11. See Supplementary Material on Science online.
12. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
13. N. Patterson *et al.*, Ancient admixture in human history. *Genetics*. **192**, 1065–1093 (2012).
14. M. Rasmussen *et al.*, The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Science*. **506**, 225–229 (2014).
15. C. Der Sarkissian *et al.*, Evolutionary Genomics and Conservation of the Endangered Przewalski’s Horse. *Curr. Biol.* **25**, 2577–2583 (2015).
16. K. Leppälä, S. V. Nielsen, T. Mailund, admixturegraph: an R package for admixture graph manipulation and fitting. *Bioinformatics*. **33**, 1738–1740 (2017).
17. R. R. Bellone *et al.*, Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *PLoS One*. **8**, e78280 (2013).
18. H. M. Holl *et al.*, Variant in the RFWD3 gene associated with PATN1, a modifier of leopard complex spotting. *Anim. Genet.* **47**, 91–101 (2016).

19. A. Ludwig *et al.*, Twenty-five thousand years of fluctuating selection on leopard complex spotting and congenital night blindness in horses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130386 (2015).
20. M. Schubert *et al.*, Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5661–9 (2014).
21. P. Librado *et al.*, Ancient genomic changes associated with domestication of the horse. *Science*. **356**, 442–445 (2017).
22. G. Larson, D. Q. Fuller, The Evolution of Animal Domestication. *Annu. Rev. Ecol. Evol. Syst.* **45**, 115–136 (2014).
23. B. S. Arbuckle, in *A Companion to the Archaeology of the Ancient Near East.*, P. D.T., Ed. (Oxford: Wiley-Blackwell, 2012), pp. 201–219.
24. J. Lira *et al.*, Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses. *Mol. Ecol.* **19**, 64–78 (2010).
25. E. T. Shev, The introduction of the domesticated horse in Southwest Asia. *Arc. Ethnol. Anthropol. Eurasia*. **44**, 123–136 (2016).
26. M. E. Allentoft *et al.*, Population genomics of Bronze Age Eurasia. *Nature*. **522**, 167–172 (2015).
27. C. B. Ramsey, Bayesian analysis of radiocarbon dates. *Radiocarbon*. **51**, 337–360 (2009).
28. S. Olsen, B. Bradley, D. Maki, A. Outram, in *Beyond the steppe and the sown: proceedings of the 2002 University of Chicago conference on Eurasian archaeology* (2006), pp. 89–111.
29. V. K. Merz, I. V. Merz, in *Exploration work on the Borly settlement 4*. S.Zh. Rakhimzhanova, Ed. (Margulanovsky Readings, Astana, 2012), pp. 206–211.

30. S. V. Svyatko, I. V. Mertz, P. J. Reimer, Freshwater Reservoir Effect on Re-Dating of Eurasian Steppe Cultures: First Results for Eneolithic and Early Bronze Age Northeast Kazakhstan. *Antiquity*. **76**, 77–85 (2015).
31. L. Bakker, *Haunstetten: Archäologie eines Augsburger Vororts; zehn Jahre Ausgrabungen von 1986 bis 1996* (1997).
32. P. W. Stockhammer *et al.*, Rewriting the Central European Early Bronze Age Chronology: Evidence from Large-Scale Radiocarbon Dating. *PLoS One*. **10**, e0139705 (2015).
33. J. P. Mallory, D. Q. Adams, *Encyclopedia of Indo-European Culture* (Taylor & Francis, 1997).
34. C. Keyser *et al.*, Ancient DNA provides new insights into the history of south Siberian Kurgan people. *Hum. Genet.* **126**, 395–410 (2009).
35. W. Fitzhugh, “The Mongolian deer stone-khirigsuur complex: dating and organization of a Late Bronze Age menagerie” in *Current Archaeological Research in Mongolia*, J. Bemmman, H. Parzinger, E. Pohl, D. Tseveendorzh, Eds. (Bonn Contributions to Asian Archaeology, Bonn, 2009), pp.183–199.
36. B. Miller, F. Allard, D. Erdenebaatar, C. Lee, A Xiongnu tomb complex: Excavations at Gol Mod 2 Cemetery, Mongolia (2002--2005). *Mongolian J. Anthropol. Arc. Ethnol.* **2**, 1–21 (2006).
37. S. Bökönyi, Die frühalluviale Wirbeltierfauna Ungarns (vom Neolithikum bis zur La Tène Zeit). *Acta Archaeol Hung.* **11**, 39–102 (1959).
38. M. L. S. Sørensen, K. Rebay-Salisbury, Landscapes of the body: Burials of the Middle Bronze Age in Hungary. *European Journal of Archaeology*. **11**, 49–74 (2007).
39. I. Poroszlai, in *Hungarian Archaeology at the turn of the Millennium*, Budapest, Z. Visy, M. Nagy, L. Bartosiewicz, Eds. (Ministry of National Cultural Heritage, Budapest, 2003), pp.141-142.



40. S. Haimovici, Materialul faunistic de la Girbovat. Studiu arheozoologic. *Archeologia Moldovei*. **14**, 153–166 (1991).
41. R. H. Dyson, The Iron Age architecture at Hasanlu: an essay. *Expedition*. **31**, 107 (1989).
42. R. H. Dyson, Rediscovering Hasanlu. *Expedition*. **31**, 3 (1989).
43. M. D. Danti, in *The Oxford Handbook of Ancient Iran*, D. T. Potts, Ed. (Oxford University Press, 2013).
44. S. Hansen, *Aktuelle Forschungen in Eurasien* (Deutsches Archäologisches Institut, Berlin, 2014).
45. L. Maldre, H. Luik, “Horse in Estonia in the Late Bronze Age: archaeozoological and archaeological data” in *The Horse and Man in European Antiquity (Worldview, Burial Rites, and Military and Everyday Life)*. A. Bliujiene, Ed. (Klaipeda University Press, Klaipeda, 2009), pp 37–47.
46. V. Lang, *The Bronze and Early Iron Ages in Estonia* (University of Tartu Press, 2012).
47. L. Maldre, Karjakasvatusest Ridala pronksiaja asulas. Loodus, inimene ja tehnoloogia. **2**, 263–276 (2008).
48. V. Onar, G. Pazvant, A. Armutak, in *Istanbul Archaeological Museums, Proceedings of the 1st Symposium on Marmaray-Metro Salvage Excavations* (İstanbul Arkeoloji Müzeleri Müdürlüğü, 2010), pp. 249–256.
49. V. Onar, G. Pazvant, E. Pasicka, A. Armutak, H. Alpak, Byzantine horse skeletons of Theodosius Harbour: 2. Withers height estimation. *Rev. Med. Vet.* **166**, 30–42 (2015).
50. J. Hansman, D. Stronach, Excavations at Shahr-I Qmis, 1971. *J. R. Asiat. Soc. GB Irel.* **106**, 8–22 (1974).

51. J. Hansman, D. Stronach, A Sasanian repository at Shahr-i Qmis. *J. R. Asiat. Soc. GB Irel.* **102**, 142–155 (1970).
52. P. Librado *et al.*, Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6889–97 (2015).
53. L. Orlando *et al.*, Recalibrating Equus evolution using the genome sequence of an early Middle Peistocene horse. *Nature.* **499**, 74–78 (2013).
54. L. S. Andersson *et al.*, Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature.* **488**, 642–646 (2012).
55. H. Kim *et al.*, Peeling back the evolutionary layers of molecular mechanisms responsive to exercise-stress in the skeletal muscle of the racing horse. *DNA Res.* **20**, 287–298 (2013).
56. K.-T. Do *et al.*, Genomic characterization of the Przewalski' s horse inhabiting Mongolian steppe by whole genome re-sequencing. *Livest. Sci.* **167**, 86–91 (2014).
57. K. Jäderkvist *et al.*, The DMRT3 “Gait keeper” mutation affects performance of Nordic and Standardbred trotters. *J. Anim. Sci.* **92**, 4279–4286 (2014).
58. J. Jun *et al.*, Whole genome sequence and analysis of the Marwari horse breed and its genetic origin. *BMC Genomics.* **15** Suppl 9, S4 (2014).
59. J. Metzger *et al.*, Next generation sequencing gives an insight into the characteristics of highly selected breeds versus non-breed horses in the course of domestication. *BMC Genomics.* **15**, 562 (2014).
60. C. Der Sarkissian *et al.*, Shotgun microbial profiling of fossil remains. *Mol. Ecol.* **23**, 1780–1798 (2014).
61. C. J. Finno *et al.*, SERPINB11 frameshift variant associated with novel hoof specific phenotype in Connemara ponies. *PLoS Genet.* **11**, e1005122 (2015).

62. M. Frischknecht *et al.*, A Non-Synonymous HMGA2 Variant Decreases Height in Shetland Ponies and Other Small Horses. *PLoS One*. **10**, e0140749 (2015).
63. P. A. Leegwater *et al.*, Dwarfism with joint laxity in Friesian horses is associated with a splice site mutation in B4GALT7. *BMC Genomics*. **17**, 839 (2016).
64. H. Jónsson *et al.*, Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18655–18660 (2014).
65. W. Haak *et al.*, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. **522**, 207–211 (2015).
66. S. Sherratt, in *Contact and exchange in the ancient world*, Mair V, Ed. (Univ. of Hawaii Press, Honolulu, 2006). pp. 30-61.
67. M. Jones *et al.*, Food globalization in prehistory. *World Archaeol.* **43**, 665–675 (2011).
68. A. K. Outram, in *The Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers*, V. Cummings, P. Jordan, M. Zvelebil, Eds. (Oxford University Press, 2014).
69. A. K. Outram, in *The Cambridge World History*, G. Barker, C. Goucher, G. Barker, C. Goucher, Eds. (Cambridge University Press, Cambridge, 2015), pp. 161–185.
70. C. Kremenetski, P. Tarasov, A. Cherkinsky, Postglacial development of Kazakhstan pine forests. *Geographie physique et Quaternaire*. **51**, 391–404 (1997).
71. V. F. Zaibert, *Botaiskaya Kultura*. (KazAkparat, Almaty, 2009).
72. M. Levine, in *Traces of Ancestry: studies in honour of Colin Renfrew*, M. Jones, Ed. (McDonald Institute for Archaeological Research, Cambridge, 2004), pp. 115-26.

73. C. French, M. Kousoulakou, “Geomorphological and micromorphological investigations of palaeosols, valley sediments and a sunken floored dwelling at Botai, Kazakhstan” in *Prehistoric steppe adaptation and the horse*, M. Levine, C. Renfrez, K. Boyle, Eds. (McDonald Institute, Cambridge, 2003), pp.105–114.
74. A. Outram, P. Rowley-Conwy, Meat and Marrow Utility Indices for Horse (Equus). *J. Archaeol. Sci.* **25**, 839–849 (1998).
75. D. Brown, D. Anthony, Bit Wear, Horseback Riding and the Botai Site in Kazakstan. *J. Archaeol. Sci.* **25**, 331– 347 (1998).
76. M. A. Zeder, “Pathways to animal domestication” in *Biodiversity in agriculture: Domestication, evolution and sustainability*, P. Gepts, T. R. Famula, R. L. Bettinger, S. B. Brush, A. B. Damania, P. E. McGuire, C. O. Qualset, Eds. (Cambridge University Press, 2012), pp. 227–259.
77. V. Eisenmann, Metapodials, System of measurements (2009);  
<http://www.vera-eisenmann.com/metapodials-system-of-measurements>
78. O. Bignon, Habitat préférentiel et connectivité des chevaux tardiglaciaires d’Europe occidentale (Equus caballus arcelini, Guadelli 1991). *Archaeofauna.* **14**, 267–284 (2005).
79. O. Bignon, Approche morphométrique des dents jugales déciduales d’Equus caballus arcelini (sensu lato, Guadelli 1991): critères de détermination et estimation de l’âge d’abattage. *C. R. Palevol.* **5**, 1005–1020 (2006).
80. O. Bignon, V. Eisenmann, in *Equids in time and space: papers in honour of Vera Eisenmann*, M. Mashkour, Ed. (Oxbow, Oxford, 2006), pp. 161–171.
81. O. Bignon-Lau, R. Cornette, M. Baylac, in *Archaeozoology of Gönnersdorf*, M. Street, E. Turner, Eds. (RömischGermanischen Zentralmuseums Verlag, Mainz, 2013), pp. 253-275.

82. O. Bignon-Lau, M. Lázníková-Galetová, Of horse metapodials debitage during the Upper Magdalenian in Europe: An overview of techniques, methods and operational sequences. *Quat. Int.* **403**, 68–78 (2016).
83. C. Gamba *et al.*, Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol. Ecol. Resour.* **16**, 459–469 (2016).
84. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* (2010).
85. C. M. Wade *et al.*, Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*. **326**, 865–867 (2009).
86. A. W. Briggs *et al.*, Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
87. M. Schubert *et al.*, Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).
88. X. Xu, U. Arnason, The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene*. **148**, 357–362 (1994).
89. B. Wallner *et al.*, Y Chromosome Uncovers the Recent Oriental Origin of Modern Stallions. *Curr. Biol.* **27**, 2029–2035 (2017).
90. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes*. **9**, 88 (2016).
91. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).

92. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
93. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* **29**, 1682–1684 (2013).
94. N. Rohland, E. Harney, S. Mallick, S. Nordenfelt, D. Reich, Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015).
95. A. W. Briggs *et al.*, Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14616–14621 (2007).
96. Picard tools (2017); <https://broadinstitute.github.io/picard/>
97. S. Xu *et al.*, High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. *J. Genet. Genomics.* **34**, 720–729 (2007).
98. Q. Jiang *et al.*, The complete mitochondrial genome and phylogenetic analysis of the Debao pony (*Equus caballus*). *Mol. Biol. Rep.* **38**, 593–599 (2011).
99. H. Goto *et al.*, A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. *Genome Biol. Evol.* **3**, 1096–1106 (2011).
100. S. Lippold, N. J. Matzke, M. Reissmann, M. Hofreiter, Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol. Biol.* **11**, 328 (2011).
101. A. Achilli *et al.*, Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2449–2454 (2012).

102. S. Sawyer, J. Krause, K. Guschanski, V. Savolainen, S. Pääbo, Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One*. **7**, e34131 (2012).
103. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).
104. T. M. Keane, C. J. Creevey, M. M. Pentony, T. J. Naughton, J. O. McInerney, Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* **6**, 29 (2006).
105. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
106. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
107. A. Rambaut, M. A. Suchard, D. Xie, A. J. Drummond, Tracer v1.6 (2014); <http://beast.bio.ed.ac.uk/Tracer>.
108. R Core Team, Others, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria (2017); <http://www.R-project.org>
109. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. **15**, 356 (2014).
110. S. A. Berger, D. Krompass, A. Stamatakis, Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.* **60**, 291–302 (2011).
111. H. Li, Improving SNP discovery by base alignment quality. *Bioinformatics*. **27**, 1157–1158 (2011).
112. T. S. Korneliussen, I. Moltke, NgsRelate: a software tool for estimating pairwise relatedness from nextgeneration sequencing data. *Bioinformatics*. **31**, 4009–4011 (2015).

113. F. G. Vieira, M. Fumagalli, A. Albrechtsen, R. Nielsen, Estimating inbreeding coefficients from NGS data: Impact on genotype calling and allele frequency estimation. *Genome Res.* **23**, 1852–1861 (2013).
114. G. C. Yang *et al.*, A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome (LWFS); a horse variant of Hirschsprung disease. *Hum. Mol. Genet.* **7**, 1047–1052 (1998).
115. I. D. Wijnberg *et al.*, A missense mutation in the skeletal muscle chloride channel 1 (CLCN1) as candidate causal mutation for congenital myotonia in a New Forest pony. *Neuromuscul. Disord.* **22**, 361–367 (2012).
116. R. Hauswirth *et al.*, Mutations in MITF and PAX3 Cause “Splashed White” and Other White Spotting Phenotypes in Horses. *PLoS Genet.* **8**, e1002653 (2012).
117. M. Aleman *et al.*, Association of a mutation in the ryanodine receptor 1 gene with equine malignant hyperthermia. *Muscle Nerve.* **30**, 356–365 (2004).
118. L. S. Brault, C. A. Cooper, T. R. Famula, J. D. Murray, M. C. T. Penedo, Mapping of equine cerebellar abiotrophy to ECA2 and identification of a potential causative mutation affecting expression of MUTYH. *Genomics.* **97**, 121–129 (2011).
119. S. A. Brooks, E. Bailey, Exon skipping in the KIT gene causes a Sabino spotting pattern in horses. *Mamm. Genome.* **16**, 893–902 (2005)
120. S. A. Brooks *et al.*, Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS Genet.* **6**, e1000909 (2010).



121. P. W. Christopherson, V. L. van Santen, L. Livesey, M. K. Boudreaux, A 10-base-pair deletion in the gene encoding platelet glycoprotein IIb associated with Glanzmann thrombasthenia in a horse. *J. Vet. Intern. Med.* **21**, 196–198 (2007).
122. D. Cook, S. Brooks, R. Bellone, E. Bailey, Missense mutation in exon 2 of SLC36A1 responsible for champagne dilution in horses. *PLoS Genet.* **4**, e1000195 (2008).
123. L. Y. Fox-Clipsham *et al.*, Identification of a mutation associated with fatal Foal Immunodeficiency Syndrome in the Fell and Dales pony. *PLoS Genet.* **7**, e1002133 (2011).
124. J. Gu *et al.*, Association of sequence variants in CKM (creatine kinase, muscle) and COX4I2 (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses. *Equine Vet. J. Suppl.*, 569–575 (2010).
125. M. Hansen, C. Knorr, A. J. Hall, T. E. Broad, B. Brenig, Sequence analysis of the equine SLC26A2 gene locus on chromosome 14q15 q21. *Cytogenet. Genome Res.* **118**, 55–62 (2007).
126. R. Hauswirth *et al.*, Novel variants in the KIT and PAX3 genes in horses with white-spotted coat colour phenotypes. *Anim. Genet.* **44**, 763–765 (2013).
127. E. W. Hill, J. Gu, B. A. McGivney, D. E. MacHugh, Targets of selection in the Thoroughbred genome contain 11 exercise-relevant gene SNPs associated with elite racecourse performance. *Anim. Genet.* **41**, 56–63 (2010).
128. S. Makvandi-Nejad *et al.*, Four loci explain 83% of size variation in the horse. *PLoS One.* **7**, e39929 (2012).
129. D. Mariat, S. Taourit, G. Guérin, A mutation in the MATP gene causes the cream coat colour in the horse. *Genet. Sel. Evol.* **35**, 119–133 (2003).

130. L. Marklund, M. Johansson Moller, K. Sandberg, L. Andersson, A missense mutation in the gene for melanocyte-stimulating hormone receptor (MC1R) is associated with the chestnut coat color in horses. *Mamm. Genome*. **7**, 895–899 (1996).
131. N. Orr *et al.*, Genome-wide SNP association--based localization of a dwarfism gene in Friesian dwarf horses. *Anim. Genet.* **41**, 2–7 (2010).
132. T. Révay, D. A. F. Villagómez, D. Brewer, T. Chenier, W. A. King, GTG mutation in the start codon of the androgen receptor gene in a family of horses with 64,XY disorder of sex development. *Sex Dev.* **6**, 108–116 (2012).
133. S. Rieder, S. Taourit, D. Mariat, B. Langlois, G. Guérin, Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (*Equus caballus*). *Mamm. Genome*. **12**, 450–455 (2001).
134. H. Signer-Hasler *et al.*, A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS One*. **7**, e37282 (2012).
135. F. Spirito *et al.*, Animal models for skin blistering conditions: absence of laminin 5 causes hereditary junctional mechanobullous disease in the Belgian horse. *J. Invest. Dermatol.* **119**, 684–691 (2002).
136. E. Brunberg *et al.*, A missense mutation in PMEL17 is associated with the Silver coat color in the horse. *BMC Genet.* **7**, 46 (2006).
137. S. C. Cannon, L. J. Hayward, J. Beech, R. H. Brown Jr, Sodium channel inactivation is impaired in equine hyperkalemic periodic paralysis. *J. Neurophysiol.* **73**, 1892–1899 (1995).
138. E. K. Shin, L. E. Perryman, K. Meek, A kinase-negative mutation of DNA-PK(CS) in equine SCID results in defective coding and signal joint formation. *J. Immunol.* **158**, 3565–3569 (1997).

139. R. E. Towers *et al.*, A nonsense mutation in the IKBKG gene in mares with incontinentia pigmenti. *PLoS One*. **8**, e81625 (2013).
140. R. C. Tryon, S. D. White, D. L. Bannasch, Homozygosity mapping approach identifies a missense mutation in equine cyclophilin B (PPIB) associated with HERDA in the American Quarter Horse. *Genomics*. **90**, 93–102 (2007).
141. H. J. Wagner, M. Reissmann, New polymorphism detected in the horse MC1R gene. *Anim. Genet.* **31**, 289–290 (2000).
142. E. W. Hill *et al.*, A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS One*. **5**, e8645 (2010).
143. D. I. Cruz-Dávalos *et al.*, Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.* **17**, 508–522 (2017).
144. L. S. Sandmeyer, C. B. Breaux, S. Archer, B. H. Grahn, Clinical and electroretinographic characteristics of congenital stationary night blindness in the Appaloosa and the association with the leopard complex. *Vet. Ophthalmol.* **10**, 368–375 (2007).
145. R. R. Bellone *et al.*, Association analysis of candidate SNPs in TRPM1 with leopard complex spotting (LP) and congenital stationary night blindness (CSNB) in horses. *Anim. Genet.* **41**, 207–207 (2010).
146. R. R. Bellone *et al.*, Fine-mapping and mutation analysis of TRPM1: a candidate gene for leopard complex (LP) spotting and congenital stationary night blindness in horses. *Brief. Funct. Genomics*. **9**, 193–207 (2010).
147. L. S. Sandmeyer *et al.*, Congenital stationary night blindness is associated with the leopard complex in the Miniature Horse. *Vet. Ophthalmol.* **15**, 18–22 (2012).

148. F. Imsland *et al.*, Regulatory mutations in TBX3 disrupt asymmetric hair pigmentation that underlies Dun camouflage color in horses. *Nat. Genet.* **48**, 152–158 (2016).
149. J. L. Petersen, S. J. Valberg, J. R. Mickelson, M. E. McCue, Haplotype diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fiber type proportions. *Anim. Genet.* **45**, 827–835 (2014).
150. B. Charlesworth, Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
151. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
152. T. S. Korneliussen, I. Moltke, A. Albrechtsen, R. Nielsen, Calculation of Tajima’s D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics.* **14**, 289 (2013).
153. M. Fumagalli, F. G. Vieira, T. Linderöth, R. Nielsen, ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics.* **30**, 1486–1487 (2014).
154. F. G. Vieira, F. Lassalle, T. S. Korneliussen, M. Fumagalli, Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biol. J. Linn. Soc. Lond.* **117**, 139–149 (2016).
155. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).
156. S. Holm, A Simple Sequentially Rejective Multiple Test Procedure. *Scand. Stat. Theory Appl.* **6**, 65–70 (1979).
157. R. B. Corbett-Detig, D. L. Hartl, T. B. Sackton, Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol.* **13**, e1002112 (2015).

158. S. Moon *et al.*, A Genome-wide Scan for Selective Sweeps in Racing Horses. *Asian-australas. J. Anim. Sci.* **28**, 1525–1531 (2015).
159. W. Park *et al.*, Investigation of de novo unique differentially expressed genes related to evolution in exercise response during domestication in Thoroughbred race horses. *PLoS One.* **9**, e91418 (2014).
160. L. A. F. Frantz *et al.*, Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* **47**, 1141–1148 (2015).
161. B. Gutiérrez-Gil *et al.*, Application of selection mapping to identify genomic regions associated with dairy production in sheep. *PLoS One.* **9**, e94623 (2014).
162. M. J. Montague *et al.*, Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17230–17235 (2014).
163. V. H. da Silva *et al.*, Genome-Wide Detection of CNVs and Their Association with Meat Tenderness in Nelore Cattle. *PLoS One.* **11**, e0157711 (2016).
164. A. Cagan, T. Blass, Identification of genomic variants putatively targeted by selection during dog domestication. *BMC Evol. Biol.* **16**, 10 (2016).
165. F. Zhao, S. McParland, F. Kearney, L. Du, D. P. Berry, Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet. Sel. Evol.* **47**, 49 (2015).
166. A. Esteve-Codina *et al.*, Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. *BMC Genomics.* **14**, 148 (2013).
167. D.-H. Shin *et al.*, Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. *BMC Genomics.* **15**, 240 (2014).
168. L. Xu *et al.*, Systematic Profiling of Short Tandem Repeats in the Cattle Genome. *Genome Biol. Evol.* **9**, 20–31 13 (2017).

169. H. R. Ramey *et al.*, Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics*. **14**, 382 (2013).
170. C. Wang *et al.*, Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in Chinese domestic pigs. *Mol. Ecol. Resour.* **15**, 414–424 (2015).
171. Z. Liu *et al.*, Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. *BMC Genomics*. **17**, 863 (2016).
172. F. Berg, S. Stern, K. Andersson, L. Andersson, M. Moller, Refined localization of the FAT1 quantitative trait locus on pig chromosome 4 by marker-assisted backcrossing. *BMC Genet.* **7**, 17 (2006).
173. B. Zhang, S. Kirov, J. Snoddy, WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **33**, W741–8 (2005).
174. F. Bastian *et al.*, in *Data Integration in the Life Sciences*, A. Bairoch, S. Cohen-Boulakia, C. Froidevaux, Eds. (Springer Berlin Heidelberg, 2008), pp. 124–131.
175. M. Schubert *et al.*, Zonkey: a simple, accurate and sensitive pipeline to genetically identify equine F1-hybrids in archaeological assemblages. *J. Arc. Sci.* (2017).

## Acknowledgments

We thank the British Institute of Persian Studies (BIPS) in Tehran and the National Museum of Iran for providing access to the material from Iran; the Archaeometry Laboratory of the University of Tehran; the staff of the Danish National High-Throughput DNA Sequencing Center; Cristina Gamba and Christian McCrory Constantz for technical support and/or discussions; Laurent Frantz, Dan Bradley and Greger Larson for critical reading of the manuscript. **Funding:** This work was supported by the Danish Council for Independent Research, Natural Sciences (4002-00152B); the Danish National Research Foundation (DNRF94); Initiative d'Excellence Chaires d'attractivité, Université de Toulouse (OURASI); the

Publishing in Elite Journals Program (PEJP-17), Vice Rectorate for Graduate Studies and Scientific Research, King Saud University; the Villum Fonden miGENEPI research project; the European Research Council (ERC-CoG-2015-681605); the Taylor Family-Asia Foundation Endowed Chair in Ecology and Conservation Biology; the Innovation Fund of the Austrian Academy of Sciences (ÖAW); the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management, and; the Russian Science Foundation (16-18-10265-RNF). **Author Contributions:** LO conceived the project and designed research; IJO, VZ and AKO designed and carried out field archaeological work; CG, AF, and NK performed ancient DNA laboratory work, with input from LO; PL and LO designed and coordinated computational analyses; KH, PL, MS, AA and LO performed computational analyses; NBe, KM, PWS, VP, AK, GM, NBa, LL, VO, JK, BB, SU, DE, SL, MM, HD, AM, AL, VM, VZ and LO provided/collected samples; OB, SA, AHA, KASAR, EW and LO provided reagents, measurements and material; CG, AF, KH and LO prepared figures and tables, with input from LO; CG, AF, KH, PL, IJO, AKO and LO wrote the supplementary information; AKO and LO wrote the paper, with input from all other co-authors. **Competing interests:** The authors declare that they have no competing interests. **Data and material availability:** Individual genome sequence data are available at the European Nucleotide Archive (accession no. PRJEB22390).

**Fig. 1. Sample location and corral enclosure at Botai.** (A) Archaeological sites. The number of genomes sequenced per site is reported between parentheses if greater than one. Triangles refer to the ancient genomes characterized here while diamonds indicate those previously published. Blue refers to wild ancient individuals, light and dark green refer to the first domestic clade (Botai and Borly4) and yellow to individuals of the second domestic clade (DOM2). The Botai culture site of Krasnyi Yar is indicated although no samples were analyzed from this site. (B). Magnetic gradient survey and excavation at Botai, with interpretation. The enclosure and its excavated boundary are indicated by red and yellow squares, respectively. Round black circles correspond to pit houses.

**Fig. 2. Horse genetic affinities.** (A) Principal Component Analysis of the genome variation present in 88 ancient and modern genomes. Only the first two principal components are shown. (B) Phylogenetic relationships. The tree was reconstructed on the basis of pairwise distances calculated with ~14.1 million transversion sites. Node support derive from 100 bootstrap pseudo-replicates. The archaeological site and age (ya) of ancient specimens are indicated in the first and last fields of the sample name. (C)  $f_3$ -outgroup statistics showing the pairwise genetic affinities.

**Fig. 3. Admixture graphs.** Panels (A-F) present the six scenarios tested. Panel A received decisive Bayes Factor support, as indicated below each corresponding alternative scenario tested. Domestic-Ancient and Domestic-A/B refer to three phylogenetic clusters identified within DOM2 (excluding Duk2): ancient individuals; modern Mongolian, Yakutian (including Tumeski\_CGG101397) and Jeju horses, and; all remaining modern breeds. (G) Posterior distributions of admixture proportions along  $p_1$  and  $p_2$  branches.

**Fig. 4. Phenotypic and genomic changes associated with ferality.** (A) Indices of the robustness of the 3<sup>rd</sup> metacarpal bone in various horse populations. Bd = Breadth at the middle of the diaphysis. GL = Maximal/Greatest length. Kent and Kumkeshu/Kozhai represent populations of Kazakhstan from the Iron Age and Eneolithic (Tersek culture), respectively. (B) Genotyping information at the *TRPM1* locus (chr1) and the *PATN1* modifier (chr3) for Botai/Borly4 horses. The absence, heterozygosis and homozygosis of alleles strongly associated with leopard spotting are depicted in white, dark grey and red, respectively. Crosses indicate insufficient data. The causative LTR insertion at the *TRPM1* locus is indicated by the number of reads overlapping both flanks of the insertion site. (C) Individual-based genetic loads. The mauve circle shows the PH specimen from the 19<sup>th</sup> century.



## **Supplementary Materials**

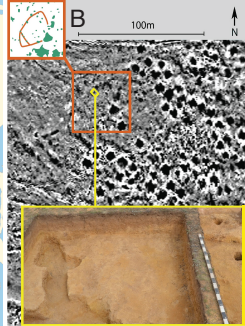
[www.sciencemag.org](http://www.sciencemag.org)

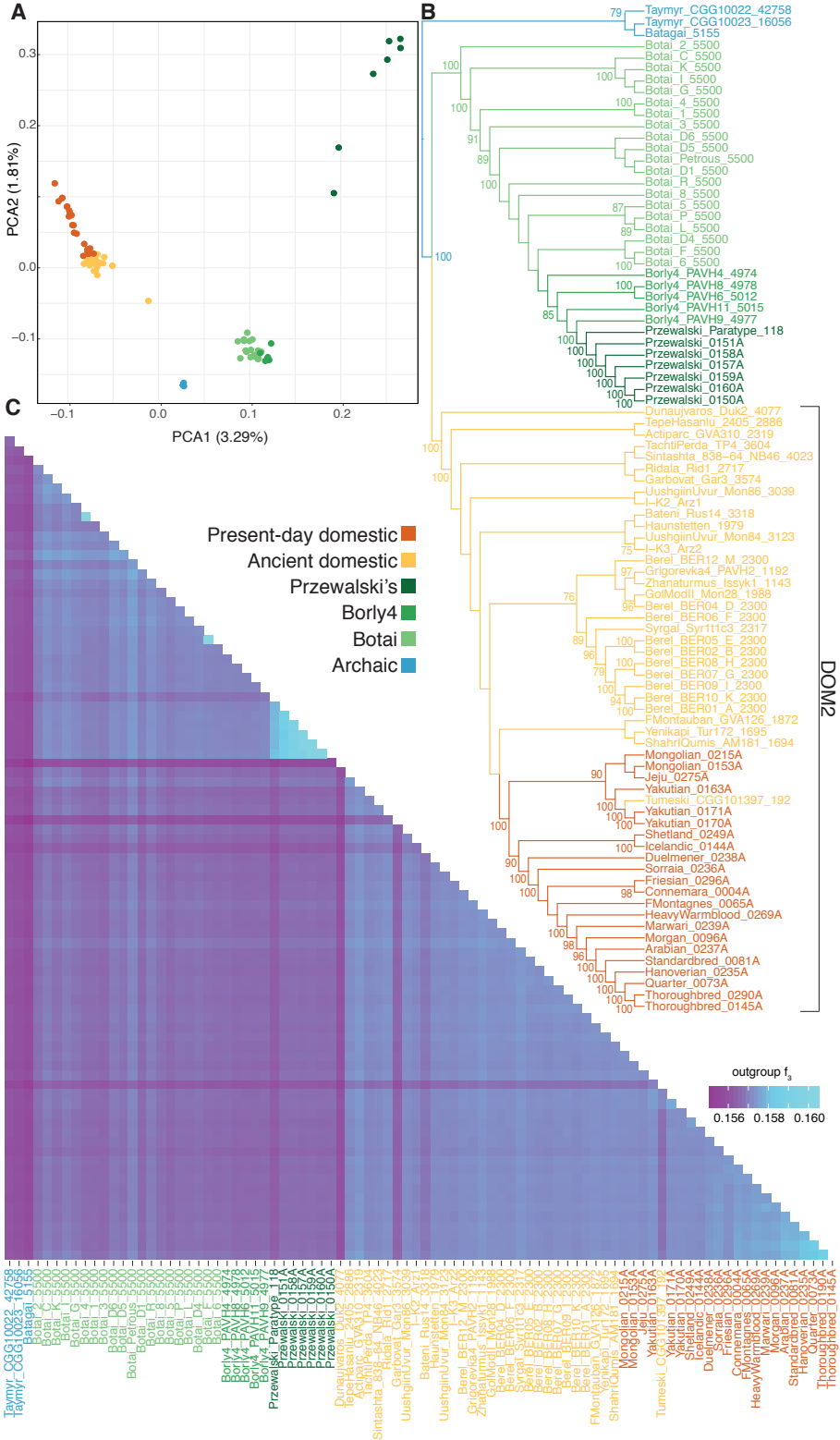
Materials and Methods

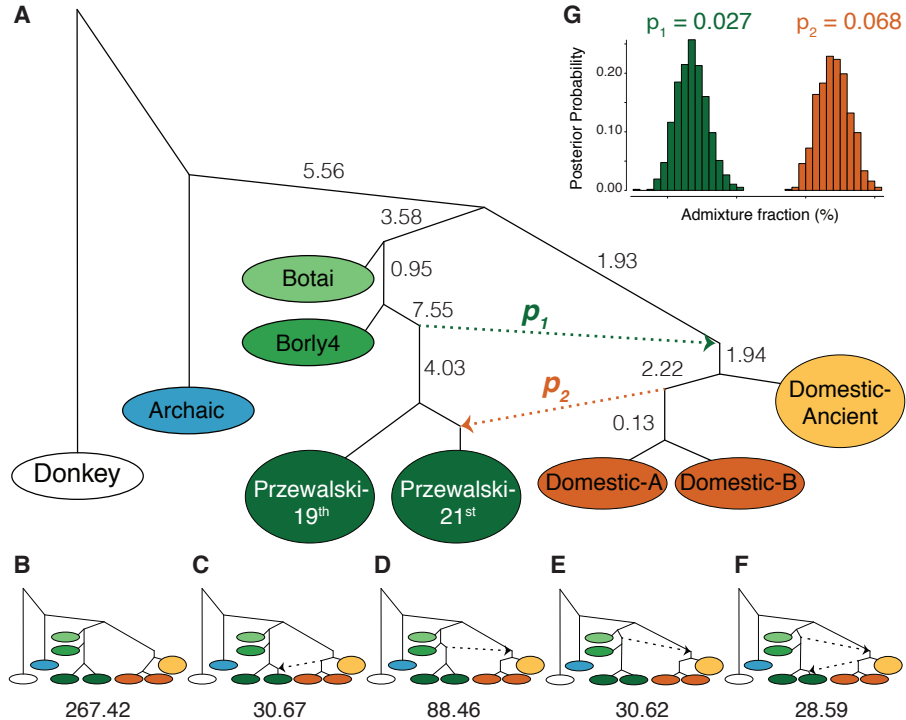
Figures S1 to S34

Tables S1 to S7 (Tables S8 to S18 as separate Excel files)

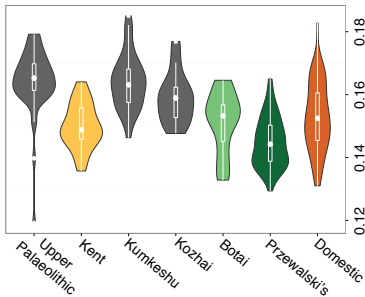
References (27-175)



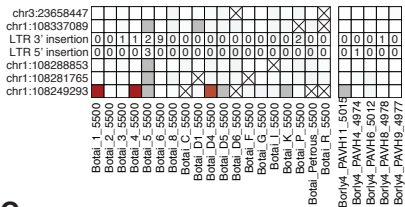




Bd/GL



# B



**C**

